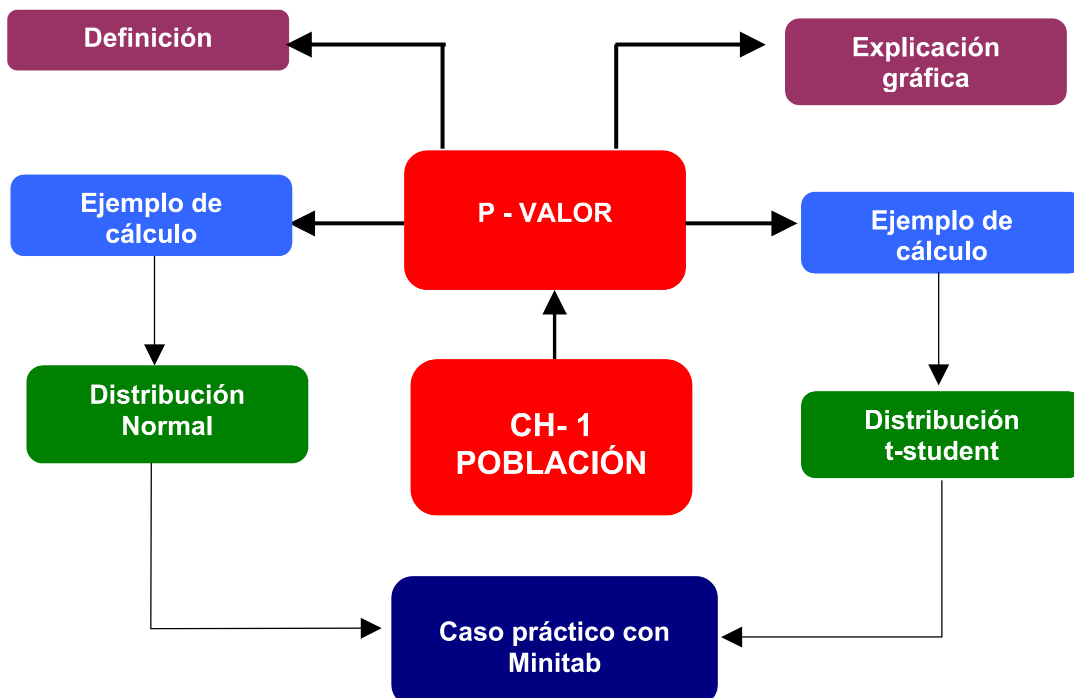
CONTRASTES DE HIPÓTESIS DE 1 POBLACIÓN

Autores: Alicia Vila (avilag@uoc.edu), Máximo Sedano (msedanoh@uoc.edu), Ángel A. Juan (ajuanp@uoc.edu), Anna López (alopezrat@uoc.edu).

ESQUEMA DE CONTENIDOS



INTRODUCCIÓN

En este *math-block*, se pretende entender qué es y para qué se utiliza un contraste de hipótesis, así como saber calcular e interpretar los p-valor a la hora de realizar dichos contrastes para la media poblacional, sea o no conocida la desviación estándar poblacional.

OBJETIVOS

- Definir una hipótesis y realizar la prueba de ésta
- Entender el concepto de p-valor
- Saber calcular el p-valor en los contrastes de hipótesis unilaterales y bilaterales
- Saber interpretar el resultado del p-valor a la hora de tomar decisiones en los contrastes de hipótesis

CONOCIMIENTOS PREVIOS

Es recomendable haber revisado los *math-blocks* correspondientes a “La distribución normal” y “Estimación puntual y intervalos de confianza”, así como los ejercicios asociados resueltos con Minitab.

CONCEPTOS FUNDAMENTALES

Concepto de contraste de hipótesis

Podemos definir un **contraste de hipótesis** como un procedimiento que se basa en lo observado en las muestras y en la teoría de la probabilidad para determinar si la hipótesis es un enunciado razonable.

Contraste de hipótesis de una población

Un **contraste de hipótesis** es un proceso estadístico que permite elegir una hipótesis de trabajo de entre dos posibles y antagónicas. El contraste comienza con la formulación de dos hipótesis sobre el valor de algún parámetro poblacional, siendo ambas incompatibles (si una es cierta, la otra necesariamente ha de ser falsa). Supondremos cierta una de ellas, a la cual llamaremos **hipótesis nula** $\mathbf{H_0}$, y trataremos de determinar hasta qué grado las observaciones registradas son coherentes con H_0. Sólo en caso de que haya fuertes indicios de incompatibilidad entre el supuesto de que H_0 sea cierta y los datos obtenidos empíricamente, descartaremos H_0 como hipótesis de trabajo y en su lugar tomaremos como cierta la **hipótesis alternativa** $\mathbf{H_1}$. Dos ejemplos de contrastes de hipótesis serían:

$$(i)\ \begin{cases} H_0 : \mu = 0 \\ H_1 : \mu \neq 0 \end{cases} \qquad (ii)\ \begin{cases} H_0 : \sigma = 2{,}5 \quad (\leq) \\ H_1 : \sigma > 2{,}5 \end{cases}$$

Contraste Bilateral (≠) Contraste Unilateral (>)

En el siguiente esquema se representan las cuatro combinaciones posibles (en función de la decisión que tomemos y de la certeza o no de la hipótesis nula) de todo contraste de hipótesis:

Decisión tomada	**Hipótesis Nula H_0** Verdadera	Falsa
No descartar H_0	Decisión correcta de tipo A **Probabilidad 1-α**	Error de tipo II **Probabilidad β**
Descartar H_0	Error de tipo I **Probabilidad α**	Decisión correcta de tipo B **Probabilidad 1-β**

Tendremos una **decisión correcta de tipo A** cuando hayamos optado por no descartar la hipótesis nula y resulte que ésta es cierta. Por su parte, una **decisión correcta de tipo B** ocurrirá cuando hayamos decidido descartar la hipótesis nula y resulte que ésta era falsa. Hablaremos de **error de tipo I** cuando hayamos descartado la hipótesis nula siendo ésta cierta (error que se considera como muy grave). Finalmente, acontecerá un **error de tipo II** cuando hayamos optado por no descartar la hipótesis nula y resulte que ésta es falsa.

Dado que descartaremos o no la hipótesis nula a partir de muestras obtenidas (es decir, no dispondremos de información completa sobre la población), no será posible garantizar que la decisión tomada sea la correcta.

Lo que sí podremos hacer es controlar la probabilidad de cometer un error. Ahora bien, ¿cuál de ellos? En un contraste de hipótesis lo interesante es rechazar la hipótesis nula. Por lo tanto el riesgo que estoy dispuesto a asumir de "equivocarme al rechazar la H_0", error de tipo I, es el que queremos controlar. Fijémonos que a error de tipo I más pequeño más seguridad al rechazar la hipótesis nula. Ahora bien, al empequeñecer el error de tipo I estamos javascript:sendmail()aumentando el error de tipo II, puesto que cuanta más probabilidad de aceptar H_0 más posibilidades de que aceptemos casos donde se cumpla H_1 (error de tipo II). Usualmente el error de tipo I se fija en 0,01, 0,05 ó 0,10.

Fijado el error de tipo I para empequeñecer el error de tipo II debemos aumentar el tamaño de muestra. Ahora bien, aumentar el número de muestra no siempre es posible ya sea por falta de presupuesto o tiempo, por inviabilidad, …

Llamaremos **potencia del contraste** a la probabilidad de rechazar la hipótesis nula siendo ésta falsa. Fijémonos que, a mayor potencia, mejor contraste, puesto que podremos aceptar la hipótesis alternativa con poca probabilidad de que sea falsa.

Denotaremos por α el **nivel de significación** o probabilidad de cometer un error de tipo I, y por β la probabilidad de cometer un error de tipo II. Con lo cual, la potencia es de $1 - \beta$. Como ya hemos indicado usualmente α se fija en 0,01, 0,05 o 0,10.

Notamos otra vez que α, β, y el tamaño muestral n están interrelacionados, de forma que si hacemos disminuir cualquiera de ellos alguno de los dos restantes habrá de aumentar. Así, p.e., si queremos tomar un α menor deberemos aceptar que aumente β o bien incrementar el tamaño de la muestra n .

Finalmente, llamaremos **estadístico de contraste** a una v.a. calculada a partir de las observaciones muestrales, la cual se usa conjuntamente con un criterio de decisión (establecido a priori) para determinar si hemos de descartar o no la hipótesis nula.

- **Concepto de p-valor.**

Definimos el **p-valor** como la probabilidad de que, suponiendo cierta H_0, el estadístico de contraste tome un valor al menos tan extremo como el que se obtiene a partir de las observaciones muestrales, i.e., el p-valor es el área de la cola de la distribución (o colas si el test es bilateral) definida a partir del estadístico de contraste:

1. El p-valor sólo puede calcularse una vez tomada la muestra, obteniéndose niveles críticos distintos para cada muestra.
2. El p-valor puede interpretarse como un nivel mínimo de significación en el sentido de que niveles de significación α, iguales o superiores al p - *valor* llevarán a rechazar la hipótesis nula.

   Por tanto, cuanto menor sea el p - *valor* mayor es el grado de incompatibilidad de la muestra con H_0, lo que lleva a rechazar H_0.
3. El cálculo del p-valor no proporciona de modo sistemático una decisión entre H_0 y H_1.

Esta forma de abordar los tests, nos permite una visión más amplia, por cuanto nos da información de para qué niveles de significación puede rechazarse la hipótesis nula, y para cuales no se puede.

Para lo que sigue, tendremos en cuenta la siguiente propiedad:

Supuesto: $\overline{X}$ se distribuye según una normal.

Recordatorio TCL: Si X se distribuye normalmente → $\overline{X}$ también lo hará. En caso contrario, necesitaremos tomar un tamaño muestral n "grande" (generalmente, $n > 30$ es suficiente).

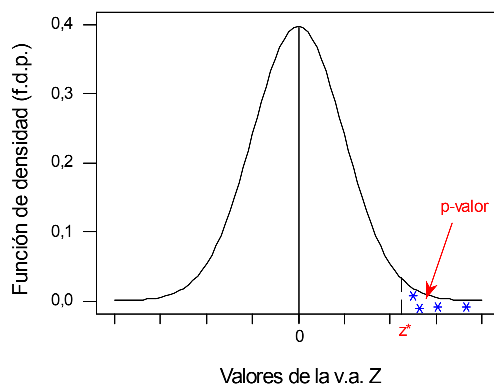
Uso del p-valor en los contrastes sobre μ con σ conocida

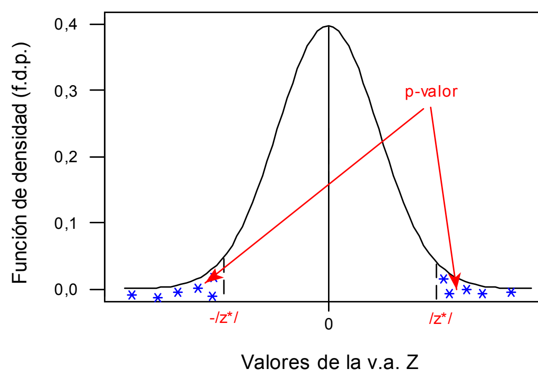
Dada una población X (que sigue una distribución cualquiera), con media μ (desconocida) y desviación estándar σ conocida, se trata de contrastar alguno de los tres tests siguientes:

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases} \quad \text{o bien} \quad \begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu < \mu_0 \end{cases} \quad \text{o bien} \quad \begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu > \mu_0 \end{cases}$$

Estadístico de contraste: $$z^* = \frac{\overline{x} - \mu}{\sigma / \sqrt{n}} \approx N(0,1)$$

- Si H_1 contiene " >" ⇒ *p-valor = P(Z>z*)*
- Si H_1 contiene " <" ⇒ *p-valor = P(Z<z*)*
- Si H_1 contiene " ≠" ⇒ *p-valor = P(Z< -/z*/ ó Z>/z*/) = 2 P(Z< /z*/)*





El p-valor nos proporciona el grado de credibilidad de la hipótesis nula: si el valor de *p* fuese "muy pequeño" (inferior a 0,001), significaría que la hipótesis nula es del todo increíble (en base a las observaciones obtenidas), y por tanto la descartaríamos; si el valor de *p* oscilase entre 0,05 y 0,001 significaría que hay fuertes evidencias en contra de la hipótesis nula, por lo que la rechazaríamos o no en función del valor que hubiésemos asignado (a priori) a α. Finalmente, si el valor de *p* es "grande" (superior a 0,05), no habría motivos suficientes como para descartar la hipótesis nula, por lo que la tomaríamos como cierta.

Criterio de decisión: Descartaremos H_0 si **p-valor** $\leq \alpha$ (normalmente α = 0,05).
En caso contrario aceptaremos H_0 (p-valor > α)

Ejemplo utilizando la tabla de la normal.

Un banco quiere analizar si las comisiones que cobra a sus clientes por operaciones en el mercado bursátil difieren significativamente de las que cobra la competencia, cuya media es de 12 euros mensuales con una desviación estándar de 4,3 euros.
Este banco toma una muestra de 64 operaciones bursátiles y observa que la comisión promedio es de 13,6 euros.
Contrastar, al nivel de significación del 5%, que este banco no difiere significativamente en el cobro de las comisiones por operaciones en la Bolsa con respecto a la competencia.

Sea X = "Comisiones que se cobran por operaciones en el mercado bursátil"
Tenemos: $X \approx (\mu,\ 4{,}3)$

Queremos contrastar:

$$H_0 : \mu = 12$$
$$H_{1:} : \mu \neq 12$$

Es decir, queremos contrastar si μ es 12 euros como la competencia o si por el contrario es distinto de esta cantidad.

Calculamos el estadístico de contraste,

$$Z^* = \frac{\overline{X} - \mu_{H_0}}{\sigma_{\overline{X}}} = \frac{\overline{X} - \mu_{H_0}}{\sigma / \sqrt{n}} = \frac{13{,}6 - 12}{4{,}3 / \sqrt{64}} = \frac{1{,}6}{0{,}5375} = 2{,}98$$

Como es un contraste de dos extremos, ahora tenemos que calcular el p-valor correspondiente a z*=2,98, es decir el área que hay por debajo de z=-2,98 más el área que hay por encima de z= 2,98, i.e., el área en las dos colas.

Si observamos la tabla de la distribución normal estándar, podemos comprobar que el área que hay a la izquierda de z=-2,98 es 0,0014 y el área que hay a la derecha de 2,98 es también 1-0,9986=0,0014 por lo que el p-valor= 2*0,0014=0,0028

Como el p-valor es menor que el nivel de significación, rechazaremos la hipótesis nula a un nivel de significación del 5%.

Por lo tanto existe evidencia estadística de que la comisión promedio que cobra este banco difiere significativamente de la competencia.

- **Uso del p-valor en los contrastes sobre μ con σ desconocida**

Dada una población X (que sigue una distribución cualquiera), con media μ y desviación estándar σ desconocidas, se trata de contrastar alguno de los tres tests siguientes:

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases} \quad \text{o bien} \quad \begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu < \mu_0 \end{cases} \quad \text{o bien} \quad \begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu > \mu_0 \end{cases}$$

Estadístico de contraste: $$t^* = \frac{\bar{x} - \mu}{s/\sqrt{n}} \approx t - Student(n-1)$$

Criterio de decisión: Descartaremos H_0 si **p-valor** $\leq \alpha$ (normalmente $\alpha = 0{,}05$).

Ejemplo utilizando la tabla de la t-student

La directora del departamento de personal de una importante corporación está reclutando un gran número de empleados para un puesto en el extranjero. Durante el proceso de selección, la administración le pregunta cómo van las cosas, y ella responde que cree que la puntuación promedio en la prueba de aptitudes será de aproximadamente 90 puntos. Cuando la administración revisa 19 de los resultados de la prueba compilados, encuentra que la puntuación media es 83,24 y la desviación estándar de esta puntuación es 11. Si la administración desea probar la hipótesis $H_0 : \mu = 90$ vs $H_a : \mu \neq 90$ al nivel de significación del 10%, ¿Cuál es el valor del estadístico de contraste y su p-valor?

$$H_0 : \mu = 90$$
$$H_a : \mu \neq 90$$

Suponemos que la población de resultados de todos los candidatos sigue una distribución normal . $X \approx N(\mu;\sigma)$ y entonces la distribución muestral de cada media muestral de cada muestra de cada población seguirá también una normal :

$$\bar{X} \approx N\left(\mu;\ \frac{S}{\sqrt{n}}\right)$$

Como no se conocen las desviaciones estándar de las dos poblaciones, tendremos que utilizar la distribución de la t-student como distribución del estadístico de contraste .

$$t = \frac{\bar{X} - \mu_{H_0}}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu_{H_0}}{\sigma/\sqrt{n}} \approx t - student(n-1)$$

Si calculamos el estadístico t de contraste nos queda:

$$t = \frac{\overline{X} - \mu_{H_0}}{\hat{\sigma}_{\overline{X}}} = \frac{\overline{X} - \mu_{H_0}}{\hat{\sigma}/\sqrt{n}} = \frac{83{,}25 - 90}{11/\sqrt{19}} = -2{,}6747$$

Como los grados de libertad son 18, entonces como tenemos un contraste de dos colas, es decir en la hipótesis alternativa aparece el distinto, es decir $H_0 : \mu = 90 \quad H_1 : \mu \neq 90$; entonces el p-valor de t = -2,6747 será la probabilidad de estar por encima de 2,6747 más la probabilidad de estar por debajo de t =-2,6747. Cuando no aparece en la tabla de la t-student el valor exacto del estadístico del cual se quiere calcular su p-valor, se toma como referencia el valor más cercano, en este caso t=-2,5524. Por tanto el p-valor = P(t>2,5524)+P(t<-2,5524)=0,01+0,01=2*0,01=0,02, porque a la derecha de 2,5524 hay la misma probabilidad que a la izquierda de -2,5524 Así que el p-valor de t=-2,6747 será menor a 0,02 porque a mayor valor del estadístico menor área por encima como se puede ver en la tabla.

Cuando los grados de libertad no aparezcan en la tabla de la t-student, se toma los grados de libertad más cercanos al cual se quiere tener en cuenta.

Si el contraste hubiese sido de una cola, bien por la derecha o bien por la izquierda, $H_1 : \mu > 90$ ó $H_1 : \mu < 90$, entonces el p–valor del estadístico (supongamos que el estadístico es t = 2,6747) si el contraste es de cola derecha, es decir (mayor que), sería la probabilidad de estar por encima de t = 2,5524 que sería 0,01, por lo que el p-valor de t= 2,6747 sería menor que 0,01.

Si es por la cola izquierda (es decir menor que), el p-valor del estadístico (supongamos que el estadístico vale t= -2,6747) sería la probabilidad de estar por debajo de t = -2,5524 que sería 0,01, por lo que el p-valor de t= - 2,6747 sería menor que 0,01.

- **Uso del p-valor en los contrastes sobre la prob. de éxito p en una binomial**

Supongamos que una población *X* se distribuye según una binomial con probabilidad de éxito *p* desconocida. A fin de estimar dicho parámetro, tomamos una muestra de tamaño *n* y definimos la **probabilidad muestral de éxito** como: *p' = nº éxitos observados / n* . Se tratará de contrastar alguno de los tres tests siguientes:

$$\begin{cases} H_0 : p = p_0 \\ H_1 : p \neq p_0 \end{cases} \quad \text{o bien} \quad \begin{cases} H_0 : p = p_0 \\ H_1 : p < p_0 \end{cases} \quad \text{o bien} \quad \begin{cases} H_0 : p = p_0 \\ H_1 : p > p_0 \end{cases}$$

Supuesto 1: La distribución de *X* es aproximadamente normal.
Recordemos que si $n \geq 20$, $n*p \geq 5$, y $n*(1-p) \geq 5$, entonces $X \approx N\left(np, \sqrt{np(1-p)}\right)$.

Supuesto 2: Las *n* observaciones que constituyen la muestra han sido seleccionadas de forma aleatoria e independiente de una población que no ha cambiado durante el muestreo.

Estadístico de contraste: $$z^* = \frac{p' - p}{\sqrt{\dfrac{p(1-p)}{n}}} \approx N(0,1)$$

Criterio de decisión: Descartaremos H_0 si **p-valor** $\leq \alpha$ (normalmente $\alpha = 0{,}05$).

Ejemplo utilizando la tabla de la normal

Un portal e-business sabe que el 60% de todos sus visitantes a la web están interesados en adquirir sus productos pero son reacios al comercio electrónico y no realizan finalmente la compra vía Internet. Sin embargo, en la dirección del portal se piensa que en el último año, el porcentaje de gente que está dispuesta a comprar por Internet ha aumentado y eso se debe reflejar en sus resultados empresariales. Contrastar al nivel de significación del 2% si en el último año se ha reducido el porcentaje de gente que no está dispuesta a comprar por Internet, si para ello se tomó una muestra de 500 visitantes para conocer su opinión y se observó que el 55% no estaba dispuesto a realizar compras vía on-line.

$$H_0 : p = 0{,}6$$
$$H_{A:} : p < 0{,}6$$

La distribución del número de visitantes del portal que no están dispuestos a comprar vía internet se va a aproximar a una normal debido a que n*p = 500*0,6 = 300 ≥ 5.

Calculamos el estadístico de control estandarizado:

$$z^* = \frac{p' - p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{0{,}55 - 0{,}6}{\sqrt{\frac{0{,}6(1-0{,}6)}{500}}} = -2{,}27$$

Como es un contraste de un extremo, por la izquierda, ahora tenemos que calcular el p-valor correspondiente a z*=-2,27 es decir el área que hay bajo la curva a la izquierda de -2,27 en la tabla de la normal. Entonces si vamos a la tabla de la normal, podemos ver que el área que hay por debajo de z = -2,27 es 0,0116.

Como el p-valor = 0,0116 es menor que el nivel de significación que es 0,02(2%), entonces rechazaremos la hipótesis nula a un nivel de significación del 2%..

En conclusión existe evidencia estadística que la proporción de visitantes al portal que están dispuestos a comprar a través de Internet ha aumentado o dicho de otra manera que el porcentaje de visitantes que son reacios a comprar vía on-line ha disminuido.

Casos hipotéticos:

1. Si el contraste hubiese sido unilateral por la derecha, es decir $p > 0{,}6$ y z= 2,27, tendríamos que tener en cuenta el áreas que hay por encima de z=-2,27,, es decir 1-área por debajo de z=2,27, es decir $1 - P(Z < 2{,}27) = 1 - 0{,}9884 = 0{,}0116$

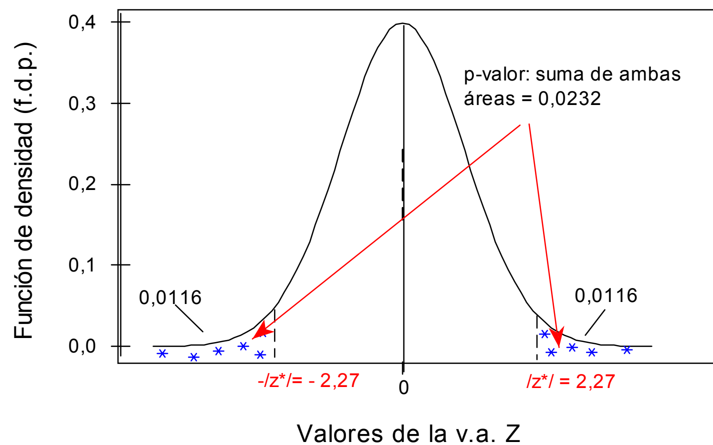
   Como el p-valor=0,0116 es menor que el nivel de significación que es 0,02(2%), entonces rechazaremos la hipótesis nula a un nivel de significación del 2%.

   En conclusión existe evidencia estadística que la proporción de visitantes al portal que están dispuestos a comprar a través de Internet ha disminuido o dicho de otra manera que el porcentaje de visitantes que son reacios a comprar vía on-line ha aumentado.

2. Si el contraste hubiese sido bilateral, es decir en la alternativa hubiese aparecido $p \neq 0{,}6$ y z = 2,27 el p-valor sería igual a la suma del área por encima de z=2,27 más el área por debajo de z = -2,27, es decir 0,0116 dos veces, p-valor = 2*0,0116=0,0232.

   Como el p-valor=0,0232 es mayor que el nivel de significación que es 0,02(2%), entonces no rechazaremos la hipótesis nula a un nivel de significación del 2%.

   En conclusión existe evidencia estadística que el porcentaje de visitantes que son reacios a comprar vía on-line es del 60%.



CASOS PRÁCTICOS CON SOFTWARE

1. Una multinacional desea analizar el sueldo neto por año de sus empleados en las empresas situadas en España. Para ello se tomó una muestra de 40 directivos y se obtuvo el salario bruto anual de cada uno. En los últimos años se había estimado que el salario medio anual de los trabajadores en España era de 18.000 euros, con una desviación estándar de 2.600 euros.

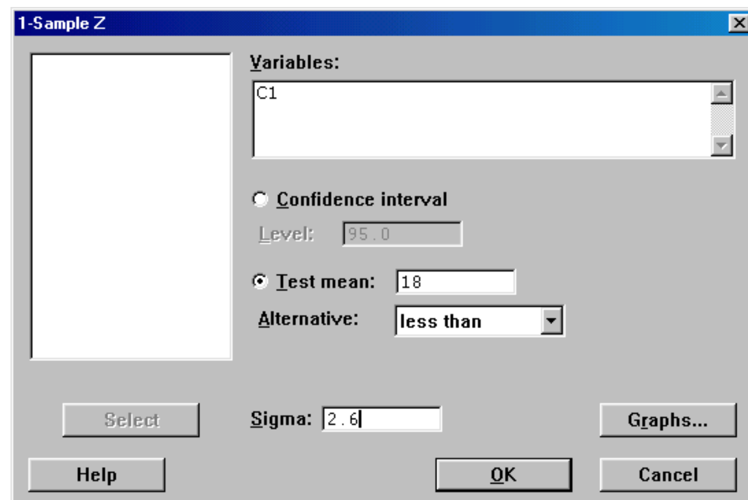
Salarios brutos medios anuales en miles de euros.

18,2630	20,3956	18,4742	18,5457	19,2611
13,2104	18,8842	18,9199	14,7056	16,6708
17,8783	20,6542	18,2433	16,8871	21,1429
15,2542	16,648	19,1553	16,1106	18,8906
20,7617	22,4938	18,2460	17,2399	17,5592
15,1401	23,5305	20,2367	16,5339	20,8730
19,.8144	14,9760	20,3570	17,5765	16,6488
16,9471	22,3734	18,2373	14,6103	12,8947

a) El jefe de personal considera que el sueldo medio anual debe ser menor que 18.000 euros y quiere contrastar, con un nivel de significación del 0.05, la hipótesis "oficial" de que el tiempo medio es de 18.000 euros frente a la hipótesis de que dicha media es menor.

El contraste de hipótesis que estableceremos será H_0: $\mu=18$ vs. H_1: $\mu<18$

Seleccionamos *Stat* > *Basic Statistics* > *1-Sample Z*:



```
Z-Test

Test of mu = 18.000 vs mu < 18.000

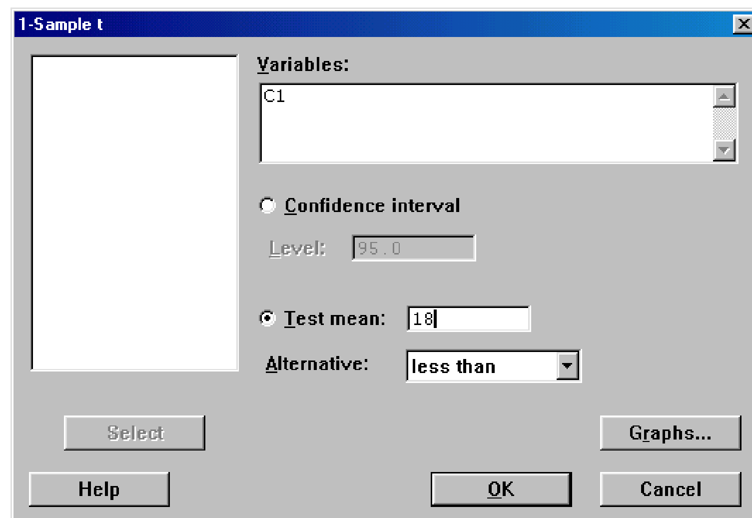
The assumed sigma = 2.60

Variable      N      Mean    StDev   SE Mean        Z          P
C1           40    18.129    2.474     0.411     0.31       0.62
```

Dado que el p-valor obtenido 0.62 > 0.05, no descartaremos la hipótesis nula, esto significa que parece razonable considerar que el salario medio bruto anual sea de es de 18.000 euros.

b) Igualmente, realiza el mismo contraste que en el apartado a), pero suponiendo ésta vez que no conoces la desviación estándar.

Análogamente, seleccionamos *Stat* > *Basic Statistics* > *1-Sample t*, obteniendo los siguientes resultados:




```
T-Test of the Mean

Test of mu = 18.000 vs mu < 18.000

Variable      N      Mean    StDev   SE Mean        T          P
C1           40    18.129    2.474     0.391     0.33       0.63
```

Por tanto, observamos que el p-valor 0.63 > 0.05, lo cual nos indica que no rechazaremos la hipótesis nula, es decir, asumiremos como posible la opción de que el salario bruto medio anual sea 18.000 euros, ya que no tenemos indicios suficientes para rechazar esta posibilidad.

2. El Dpto. de Marketing de una empresa europea quiere analizar la eficacia de su fuerza de ventas. Para ello tomó una muestra de 150 comerciales repartidos por sus varias delegaciones en Europa y se obtuvo en euros lo que cada comercial ha facturado en los últimos seis meses. Se ha comprobado que, hasta ahora, que el volumen facturado por la fuerza de ventas hasta ahora seguía una distribución aproximadamente normal de media 165.000 € y desviación típica de 45.000 €.

MINITAB - Untitled - [PPISO.MTW ***]

File Edit Manip Calc Stat Graph Editor Window Help

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11
→											
1	179000	80500	151500	223000	183000	98000	195000	121900			
2	126500	152000	141000	163500	85000	137900	97000	126000			
3	134500	231750	119000	78000	126500	108000	117000	164900			
4	125000	180000	108500	187000	162000	120500	150000	172000			
5	142000	185000	126500	133000	169000	128500	180500	100000			
6	164000	265000	302000	125000	175000	142500	160000	129900			
7	146000	135000	130000	116000	267000	72000	181500	110000			
8	129000	203000	140000	135000	150000	124900	124000	131000			
9	141900	141000	123500	194500	115000	134000	125900	107000			
10	135000	159000	153500	99500	126500	205406	165000	165900			
11	118500	182000	194900	152500	215000	217000	122000				
12	160000	208000	165000	141900	190000	94000	132000				
13	89900	96000	179900	139900	190000	189900	145900				
14	169900	156000	194500	117500	113500	168500	156000				
15	127500	185500	127500	150000	116300	133000	136000				
16	162500	275000	170000	177000	190000	180000	142000				
17	152000	144900	160000	136000	145000	139500	140000				
18	122500	155000	135000	158000	269900	210000	144900				
19	220000	110000	117000	211900	135500	126500	133000				
20	141000	154000	235000	165000	190000	285000	196800				

Current Worksheet: PPISO.MTW 12:06

Inicio est-... Hot... Ned... Uoc 2ºse... Win... SOL... MI... 12:06

a) Realizar un contraste de hipótesis bilateral sobre la media de la población para un nivel de significación α=0,05.

Tomamos como hipótesis nula H_0: μ = 165000, siendo la hipótesis alternativa H_1: $\mu \neq$ 165000

Copiamos los datos en una hoja de Minitab y seleccionamos *Stat* > *Basic Statistics* > *1-Sample Z* :

1-Sample Z
Variables: Precio
Confidence interval
Level: 95.0
Test mean: 165000
Alternative: not equal
Select
Sigma: 45000
Graphs...
Help
OK
Cancel

```
Z-Test

Test of mu = 165000 vs mu not = 165000
The assumed sigma = 45000

Variable       N      Mean    StDev   SE Mean        Z          P
Precio       150    153775    41611      3674    -3.06     0.0023
```

Observar que el p-valor obtenido es de 0,0023 < 0,05. Esto indica que deberíamos rechazar la hipótesis nula.
Por tanto, concluiremos que hay indicios suficientes para pensar que la facturación obtenida por la fuerza de ventas ha variado.

b) Realizar un contraste similar al anterior suponiendo que esta vez no conocemos la desviación estándar σ.

En este caso, tenemos que utilizar la distribución t-Student, en lugar de la normal debido a que la desviación estándar es desconocida.

Seleccionamos *Stat* > *Basic Statistics* > *1-Sample t*:

1-Sample t
Variables: Precio
Confidence interval
Level: 95.0
Test mean: 165000
Alternative: not equal
Select
Graphs...
Help
OK
Cancel

```
T-Test of the Mean

Test of mu = 165000 vs mu not = 165000

Variable      N      Mean    StDev   SE Mean        T        P
Precio      150    153775    41611      3398    -3.30   0.0012
```

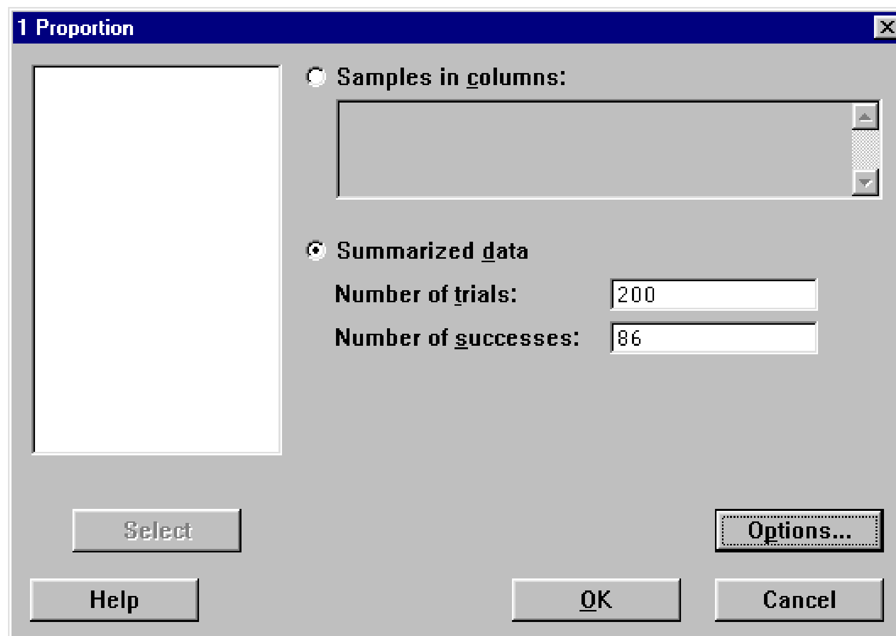
Observar que el p-valor obtenido 0,0012 sigue siendo inferior a 0,05, lo que nos lleva nuevamente a rechazar la hipótesis nula.
Por ello, podemos concluir que efectivamente la facturación obtenida por la fuerza de ventas ha variado.

3. Supongamos que trabajamos para un candidato a la alcaldía de nuestra ciudad y nos encontramos en plena campaña electoral. Nuestro candidato estima que tiene el apoyo del 55% de los votantes. Sin embargo, acaban de llegar a nuestra oficina los datos de una encuesta reciente en la que sólo 86 de 200 potenciales votantes (seleccionados de forma aleatoria) optan por nuestra opción. Nos interesa contrastar, a un nivel de significación del 0,05, las hipótesis $H_0 : p = 0{,}55$ vs. $H_1 : p < 0{,}55$.

   Observar que se verifican los supuestos. En particular, el supuesto de normalidad se verifica dado que $n = 200 > 20$, $np = 200*0{,}55 > 5$, y $n(1-p) = 200*0{,}45 > 5$.

   Realicemos el contraste:

   Seleccionamos: *Stat > Basic Statistics > 1 Proportion* :



Entramos en el menú *Options* y rellenamos los campos como se muestra en la imagen siguiente:

1 Proportion - Options

Confidence level: 95,0

Test proportion: 0,55

Alternative: less than

☑ Use test and interval based on normal distribution

Help OK Cancel

```
Test and Confidence Interval for One Proportion

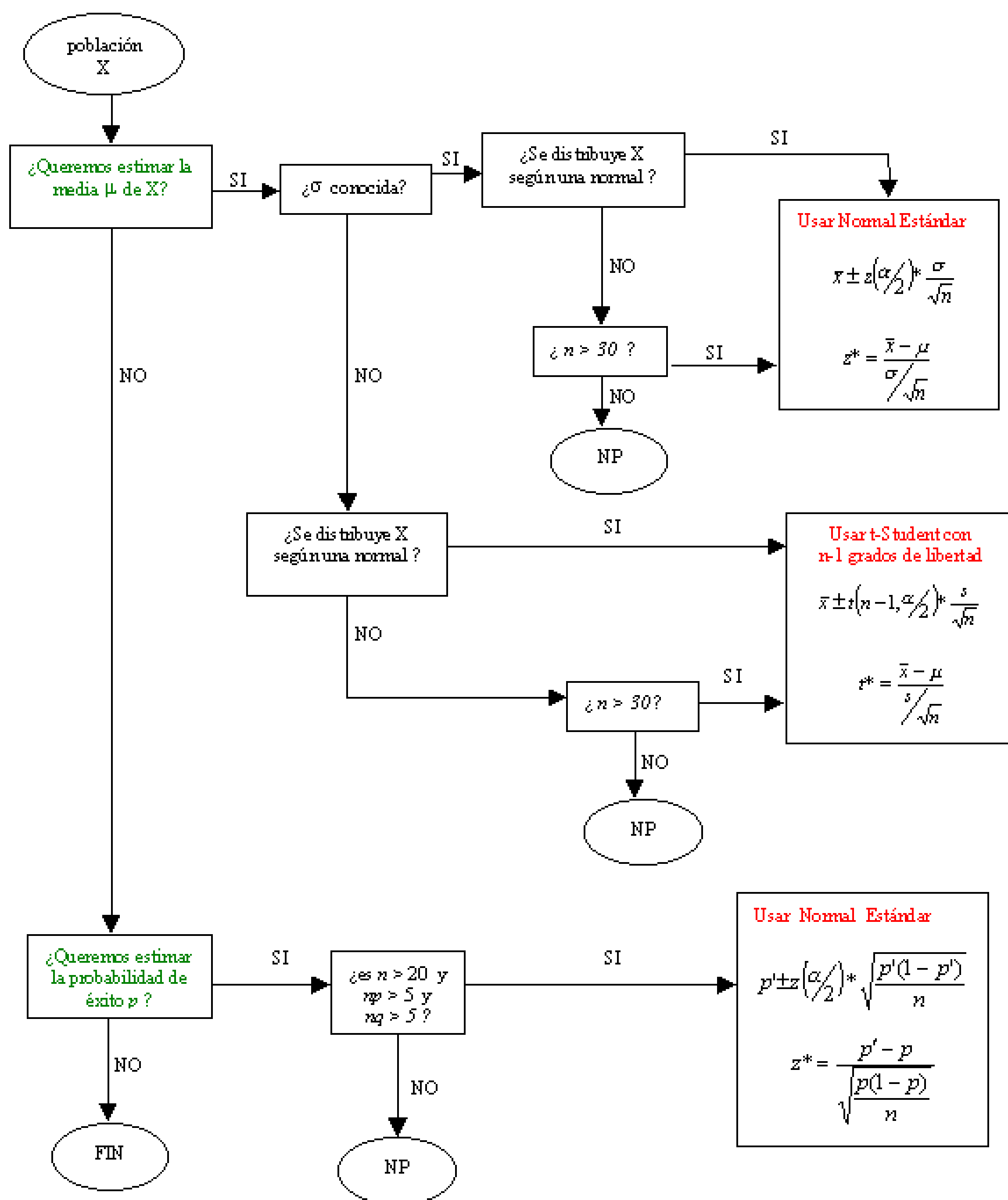
Test of p = 0,55 vs p < 0,55

Sample      X      N  Sample p         95,0 % CI          Z-Value  P-Value
1          86    200  0,430000  (0,361387; 0,498613)    -3,41    0,000
```

A raíz del resultado obtenido (p-valor = 0,000 < 0,05), concluimos que deberemos descartar la hipótesis nula, i.e., los datos obtenidos en la última encuesta sobre intención de voto sugieren que el porcentaje de votantes que apoyan nuestra candidatura es inferior al 55%.

De hecho, a partir de las observaciones, podemos afirmar con un nivel de confianza del 95% que el porcentaje de votos favorables se sitúa entre un 36% y un 50%.

RESUMEN-ESQUEMA SOBRE IC & CH PARA 1 POBLACIÓN

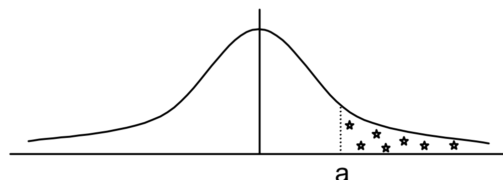


NP significa que deberemos usar métodos No Paramétricos (fuera del contenido del curso)

TABLA DE LA t-STUDENT

Esta tabla nos da los valores de a
tales que P[t(df) ≥ a] = p

donde t(df) *sigue una distribución t-Student*
con df *grados de libertad*



DF	VALORES DE P							
	0,400	0,250	0,150	0,100	0,050	0,025	0,010	0,005
1	0,3249	1,0000	1,9626	3,0777	6,3137	12,7062	31,8210	63,6559
2	0,2887	0,8165	1,3862	1,8856	2,9200	4,3027	6,9645	9,9250
3	0,2767	0,7649	1,2498	1,6377	2,3534	3,1824	4,5407	5,8408
4	0,2707	0,7407	1,1896	1,5332	2,1318	2,7765	3,7469	4,6041
5	0,2672	0,7267	1,1558	1,4759	2,0150	2,5706	3,3649	4,0321
6	0,2648	0,7176	1,1342	1,4398	1,9432	2,4469	3,1427	3,7074
7	0,2632	0,7111	1,1192	1,4149	1,8946	2,3646	2,9979	3,4995
8	0,2619	0,7064	1,1081	1,3968	1,8595	2,3060	2,8965	3,3554
9	0,2610	0,7027	1,0997	1,3830	1,8331	2,2622	2,8214	3,2498
10	0,2602	0,6998	1,0931	1,3722	1,8125	2,2281	2,7638	3,1693
11	0,2596	0,6974	1,0877	1,3634	1,7959	2,2010	2,7181	3,1058
12	0,2590	0,6955	1,0832	1,3562	1,7823	2,1788	2,6810	3,0545
13	0,2586	0,6938	1,0795	1,3502	1,7709	2,1604	2,6503	3,0123
14	0,2582	0,6924	1,0763	1,3450	1,7613	2,1448	2,6245	2,9768
15	0,2579	0,6912	1,0735	1,3406	1,7531	2,1315	2,6025	2,9467
16	0,2576	0,6901	1,0711	1,3368	1,7459	2,1199	2,5835	2,9208
17	0,2573	0,6892	1,0690	1,3334	1,7396	2,1098	2,5669	2,8982
18	0,2571	0,6884	1,0672	1,3304	1,7341	2,1009	2,5524	2,8784
19	0,2569	0,6876	1,0655	1,3277	1,7291	2,0930	2,5395	2,8609
20	0,2567	0,6870	1,0640	1,3253	1,7247	2,0860	2,5280	2,8453
21	0,2566	0,6864	1,0627	1,3232	1,7207	2,0796	2,5176	2,8314
22	0,2564	0,6858	1,0614	1,3212	1,7171	2,0739	2,5083	2,8188
23	0,2563	0,6853	1,0603	1,3195	1,7139	2,0687	2,4999	2,8073
24	0,2562	0,6848	1,0593	1,3178	1,7109	2,0639	2,4922	2,7970
25	0,2561	0,6844	1,0584	1,3163	1,7081	2,0595	2,4851	2,7874
26	0,2560	0,6840	1,0575	1,3150	1,7056	2,0555	2,4786	2,7787
27	0,2559	0,6837	1,0567	1,3137	1,7033	2,0518	2,4727	2,7707
28	0,2558	0,6834	1,0560	1,3125	1,7011	2,0484	2,4671	2,7633
29	0,2557	0,6830	1,0553	1,3114	1,6991	2,0452	2,4620	2,7564
30	0,2556	0,6828	1,0547	1,3104	1,6973	2,0423	2,4573	2,7500
35	0,2553	0,6816	1,0520	1,3062	1,6896	2,0301	2,4377	2,7238
40	0,2550	0,6807	1,0500	1,3031	1,6839	2,0211	2,4233	2,7045
45	0,2549	0,6800	1,0485	1,3007	1,6794	2,0141	2,4121	2,6896
50	0,2547	0,6794	1,0473	1,2987	1,6759	2,0086	2,4033	2,6778
60	0,2545	0,6786	1,0455	1,2958	1,6706	2,0003	2,3901	2,6603
70	0,2543	0,6780	1,0442	1,2938	1,6669	1,9944	2,3808	2,6479
80	0,2542	0,6776	1,0432	1,2922	1,6641	1,9901	2,3739	2,6387
90	0,2541	0,6772	1,0424	1,2910	1,6620	1,9867	2,3685	2,6316
100	0,2540	0,6770	1,0418	1,2901	1,6602	1,9840	2,3642	2,6259
120	0,2539	0,6765	1,0409	1,2886	1,6576	1,9799	2,3578	2,6174
150	0,2538	0,6761	1,0400	1,2872	1,6551	1,9759	2,3515	2,6090
200	0,2537	0,6757	1,0391	1,2858	1,6525	1,9719	2,3451	2,6006
300	0,2536	0,6753	1,0382	1,2844	1,6499	1,9679	2,3388	2,5923
1E+09	0,2533	0,6745	1,0364	1,2816	1,6449	1,9600	2,3263	2,5758

Financiado por la Secretaría de Estado de Educación y Universidades (MECD)

BIBLIOGRAFÍA

[1] D.A. Lind, R.D. Mason, W.G. Marchal (2001): "Estadística para Administración y Economía". Ed. Irwin McGraw-Hill.F.

[2] Kvanli, A. "Introduction to Business Statistics" South-Western

[3] R. Johnson (1996): "Elementary Statistics". Ed. Duxbury

[4] Richard I. Levin & David S. Rubin (1996): "Estadística para Administradores". Ed. Prentice Hall.

ENLACES

- http://oak.cats.ohiou.edu/~wallacd1/shyp.html : Características y ejemplos del contraste de hipótesis para una muestra, conocida la media y desviación estándar de la población.
- http://oak.cats.ohiou.edu/~wallacd1/sci.html : Características y ejemplos de intervalos de confianza para una muestra, conocida la media de la población.
- http://halweb.uc3m.es/esp/Personal/personas/stefan/ESP/applet.htm : Colección de applets de conceptos básicos de Estadística.
- http://e-stadistica.bio.ucm.es/mod_intervalos/intervalos_applet.html : Applet sobre estimación por intervalos.
- http://e-stadistica.bio.ucm.es/mod_contraste/contraste_applet.html : Applet sobre contraste de hipótesis.